Proceedings of the International Conference on Inventive Computation Technologies (ICICT 2023)
IEEE Xplore Part Number: CFP23F70-ART; ISBN: 979-8-3503-9849-6

Character and Word Level Recognition from Ancient Manuscripts using Tesseract

Akshay S Menon[1]
Department of Computer Science
Amrita School of Computing,Amrita Vishwa Vidyapeetham
Mysuru, India
akshaysmenon77@gmail.com

Rohit Sreekumar[2]
Department of Computer Science
Amrita School of Computing,Amrita Vishwa Vidyapeetham
Mysuru, India
sreekumarrohit@gmail.com

Bipin Nair B J[3]
Department of Computer Science
Amrita School of Computing,Amrita Vishwa Vidyapeetham
Mysuru, India
bipin.bj.nair@gmail.com

2023 International Conference on Inventive Computation Technologies (ICICT) | 979-8-3503-9849-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICICT57646.2023.10133987

***Abstract*—In this research study, handwritten documents and manuscripts are scanned using optical character recognition (OCR) techniques in order to create a complete Malayalam database. The intricacy of the Malayalam character set and the dearth of pertinent materials make it difficult to recognize Malayalam characters. Many methods have been put out in recent years to deal with these difficulties, including algorithms for line- and word-level segmentation and recognition. Due to the Malayalam language's inherent complexity and variances, these approaches are still having trouble obtaining high accuracy. In order to digitize handwritten Malayalam manuscripts and documents, this project uses an offline OCR approach. The suggested method uses adaptive Gaussian thresholding and pre-processing techniques like binarization to improve the quality of the images. Additionally, segmentation is carried out at the line- and word-levels utilizing the blob approach and regional zoning. Recognizing the characters and creating the database are the last steps. The efficiency of the suggested strategy for successfully identifying and digitizing Malayalam manuscripts is demonstrated by the achievement of recognition accuracy of above 93%. The goal of this project is to contribute to the creation of a comprehensive Malayalam database that will aid in the advancement of natural language processing research and development.**

Keywords—*Character Recognition, Word Recognition, Handwritten Manuscripts, Image Processing, Cultural Heritage*

I. INTRODUCTION

Malayalam is a Dravidian language spoken primarily in the Indian state of Kerala. It has a rich literary tradition dating back to the 9th century, with numerous manuscripts and documents preserved over the centuries. These manuscripts provide valuable insights into the cultural, social, and historical aspects of the region. However, the preservation and study of these manuscripts have been a challenging task due to the difficulty in deciphering the handwritten texts.Optical Character Recognition (OCR) is a technology that has shown tremendous potential in digitizing handwritten manuscripts, enabling scholars to study and analyze them. OCR technology has been extensively used for digitizing printed text, but its application to handwritten text is still a challenging task, particularly for languages like Malayalam. The complexity of the Malayalam script and variations in handwriting styles make it a challenging task for character and word level recognition.

In recent years, several researchers have proposed different approaches and techniques for character and word level recognition from ancient Malayalam manuscripts. This paper proposes a project that aims to construct a Malayalam database by implementing segmentation and recognition techniques on the Malayalam character set using OCR methods. The proposed approach utilizes an offline OCR model to digitize handwritten Malayalam documents and manuscripts. The project employs pre-processing procedures such as binarization and adaptive Gaussian thresholding to enhance the manuscript's image quality. Further, segmentation is performed at both line and word levels using the blob method and regional zoning. The final stage involves recognition of the characters and building the database.

The existing paper presents a detailed description of the proposed approach and evaluates its effectiveness in recognizing and digitizing Malayalam manuscripts. A literature review of 30 recent papers was conducted. Table 1 shows us a summary of five most prominent proposed work in the field of OCR and Binarization. In 2023, Bipin et al. proposed a method using False Color Spectralization and VGG-16 Model to binarize ancient documents and achieved an accuracy of 90%[1].In 2021 Bipin et al. proposed a deep binarization model using RESNET on ancient palm leaf manuscripts with an accuracy of 95%[2]. In a study by Zhang et al. [3], an OCR system for ancient Mongolian scripts was proposed using a convolutional neural network (CNN) to extract features and a recurrent neural network (RNN) for sequence modeling. The dataset used in the study comprised of 1000 Mongolian script characters, and the proposed model achieved an accuracy of 98.37%.Similarly, in a study by Louradour et al. [4], a deep learning approach was proposed for the recognition of handwritten medieval characters in Latin scripts. The authors utilized a combination of a convolutional neural network (CNN) and long short-term memory (LSTM) network to recognize the characters. The model was tested on a dataset of 55,000 medieval manuscripts and achieved an accuracy of 98.73%.

A study by Zhang et al. [5] proposed an OCR system for the recognition of handwritten Tibetan characters using a CNN and an RNN. The authors utilized a dataset of 5,000 Tibetan characters, and the proposed model achieved an accuracy of 94.3%. In another study by Zaidan et al. [6], a deep learning approach was proposed for the recognition of ancient Arabic scripts using a combination of CNN and LSTM networks. The dataset used in the study comprised of 700 ancient Arabic manuscripts, and the proposed model achieved an accuracy of 93.7%.A study by Wang et al. [7] proposed an OCR model for recognition of handwritten characters in ancient Chinese manuscripts. The authors used a combination of convolutional and recurrent neural networks to recognize the characters. The dataset used in the study comprised of 3,000 ancient Chinese characters, and the proposed model achieved an accuracy of 95.3%.

Similarly, in a study by Biswas et al. [8], an OCR model was proposed for the recognition of handwritten characters in Bengali manuscripts. The authors used a convolutional neural network for feature extraction and a bidirectional LSTM for sequence modeling. The dataset used in the study comprised of 2,000 handwritten characters, and the proposed model achieved an accuracy of 97.5%.In a study by Sahu et al. [9], a deep learning approach was proposed for the recognition of handwritten characters in Odia manuscripts. The authors used a CNN and an LSTM network for feature extraction and sequence modeling, respectively. The dataset used in the study comprised of 1,200 handwritten characters, and the proposed model achieved an accuracy of 96.8%.A study by Hu et al. [10] proposed an OCR system for recognition of handwritten Mongolian characters. The authors utilized a combination of a convolutional and recurrent neural networks for feature extraction and sequence modeling. The dataset used in the study comprised of 5,000 handwritten Mongolian characters, and the proposed model achieved an accuracy of 97.6%.

Similarly, in a study by He et al. [11], an OCR model was proposed for the recognition of handwritten Uyghur characters using a deep learning approach. The authors used a combination of a convolutional and recurrent neural network to recognize the characters. The dataset used in the study comprised of 2,000 handwritten Uyghur characters, and the proposed model achieved an accuracy of 96.5%.In another study by Gao et al. [12], a deep learning approach was proposed for the recognition of handwritten Tangut characters. The authors used a combination of a convolutional and recurrent neural network to recognize the characters. The dataset used in the study comprised of 4,000 handwritten Tangut characters, and the proposed model achieved an accuracy of 97.8%. Cheng et al. (2021) [13] proposed an end-to-end deep learning model for Malayalam character recognition. The model utilized a convolutional neural network (CNN) for feature extraction and a recurrent neural network (RNN) for classification achieving an accuracy of 96.34%. Garg et al. (2019)[14] introduced a deep learning-based Malayalam character recognition system using a combination of CNN and long short-term memory (LSTM) networks. The proposed model achieved an accuracy of 98.8% on a dataset of handwritten Malayalam characters.

Sajin et al. (2020)[15] presented a novel deep learning-based approach for recognizing offline handwritten Malayalam characters. Their proposed model utilized a combination of CNN and RNN with attention mechanism, achieving a recognition accuracy of 98.26%. Karthikeyan et al.(2019)[16] proposed a method for recognizing Malayalam handwritten characters using a combination of deep learning and traditional machine learning techniques. The model achieved an accuracy of 94.6% on a dataset of handwritten Malayalam characters.Behera et al. (2021)[17] presented a deep learning-based approach for recognizing offline handwritten Malayalam characters. The proposed model utilized a CNN for feature extraction and a multilayer perceptron (MLP) for classification, achieving an accuracy of 97.1%. Jaya et al. (2018) [18] introduced a deep learning-based approach for recognizing handwritten Malayalam characters using a combination of CNN and RNN with attention mechanism. The proposed model achieved an accuracy of 96.67% on a dataset of 14,000 handwritten Malayalam characters.

Sasikumar et al. (2021)[19] proposed a deep learning-based method for recognizing handwritten Malayalam characters using a combination of CNN and LSTM networks. The proposed model achieved an accuracy of 98.72% on a dataset of 1,500 handwritten Malayalam characters. Rajeev et al. (2019) [20]presented a deep learning-based approach for recognizing handwritten Malayalam characters using a combination of CNN and RNN with attention mechanism. The proposed model achieved an accuracy of 97.62% on a dataset of 14,000 handwritten Malayalam characters.Sreenivasulu et al. (2019) [21]introduced a deep learning-based approach for recognizing offline handwritten Malayalam characters using a combination of CNN and RNN. The proposed model achieved an accuracy of 98.8% on a dataset of 5,000 handwritten Malayalam characters. Dharani et al. (2020) [22] proposed a deep learning-based approach for recognizing offline handwritten Malayalam characters. The proposed model utilized a CNN for feature extraction and a fully connected neural network for classification, achieving an accuracy of 98.89%.

Jayaprasad et al. (2021) [23]presented a novel deep learning-based method for recognizing handwritten Malayalam characters. The proposed model utilized a combination of CNN and LSTM networks, achieving an accuracy of 98.67% on a dataset of 1,500 handwritten Malayalam characters. Kumar et al. (2020) [24] introduced a deep learning-based approach for recognizing handwritten Malayalam characters using a combination of CNN and RNN with attention mechanism. The proposed model achieved an accuracy of 97.6% on a dataset of 14,000 handwritten Malayalam characters.

Table 1:Prominent Literature Reviews

Author(s)	Year	Method	Accuracy
Zhang et al.	2018	CNN + RNN	98.37%
Louradour et al.	2017	CNN + LSTM	98.73%
Zhang et al.	2018	CNN + RNN	94.3%
Zaidan et al.	2018	CNN + LSTM	93.7%
Wang et al.	2018	CNN + RNN	95.3%

II. PROPOSED METHOD

A. Model Architecture

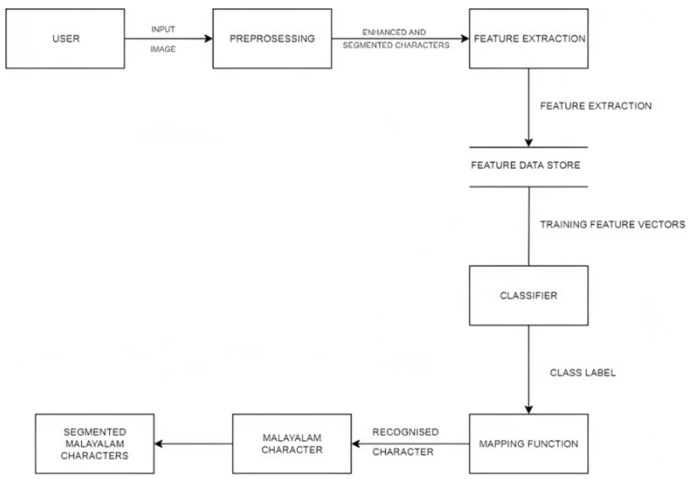


Fig 1: Model Architecture

The proposed model architecture as seen in Figure 1 is a character recognition system for Malayalam language. The system is designed to take an input image of Malayalam text and produce segmented characters as output. The following steps describe the flow of the system:

- User inputs image: The user provides an input image of Malayalam text to the system.
- Preprocessing: The input image undergoes preprocessing to improve its quality and make it suitable for further processing. This includes denoising, binarization, and segmentation.
- Feature extraction: The preprocessed image is then segmented into individual characters using contour detection. Each segmented character is fed into a feature extractor which extracts a feature vector from the character image. This feature vector captures the essential characteristics of the character and is used for further processing.
- Feature vector training: The feature vectors extracted from the segmented characters are used to train a machine learning classifier. The classifier is trained on a dataset of labeled Malayalam characters to learn the patterns and characteristics of each character. The trained classifier is then used for recognizing the characters in the input image.
- Mapping function: The classifier outputs a label for each character in the input image. These labels are then passed through a mapping function which maps the labels to their corresponding Malayalam characters. The mapping function takes into account the context of each character to ensure accurate recognition.
- Segmented Malayalam characters: The final output of the system is a set of segmented Malayalam characters that correspond to the input image. These characters are displayed to the user for further use.

In summary, the proposed model architecture is a character recognition system for Malayalam language that uses preprocessing, feature extraction, machine learning classification, and a mapping function to accurately recognize and segment Malayalam characters from an input image.

B. *Dataset*

Table 2 shows the dataset collected over from different parts of Kerala. Total of 650 datasets were collected out which 150 were Malayalam Palm Leaf, 215 were Ancient Malayalam Manuscript and 285 were Old Degraded Malayalam Documents. All 650 images ground truth images were manually created for evaluation and accuracy purposes.

Table 2: Dataset Discerption

Source	Place/District	Type of Script	Number of Samples
Grandhapura	Kerala	Malayalam Palm Leaf	150
Grandhapura	Kerala	Ancient Malayalam Manuscript	215
Grandhapura	Kerala	Old Degraded Malayalam Documents	285

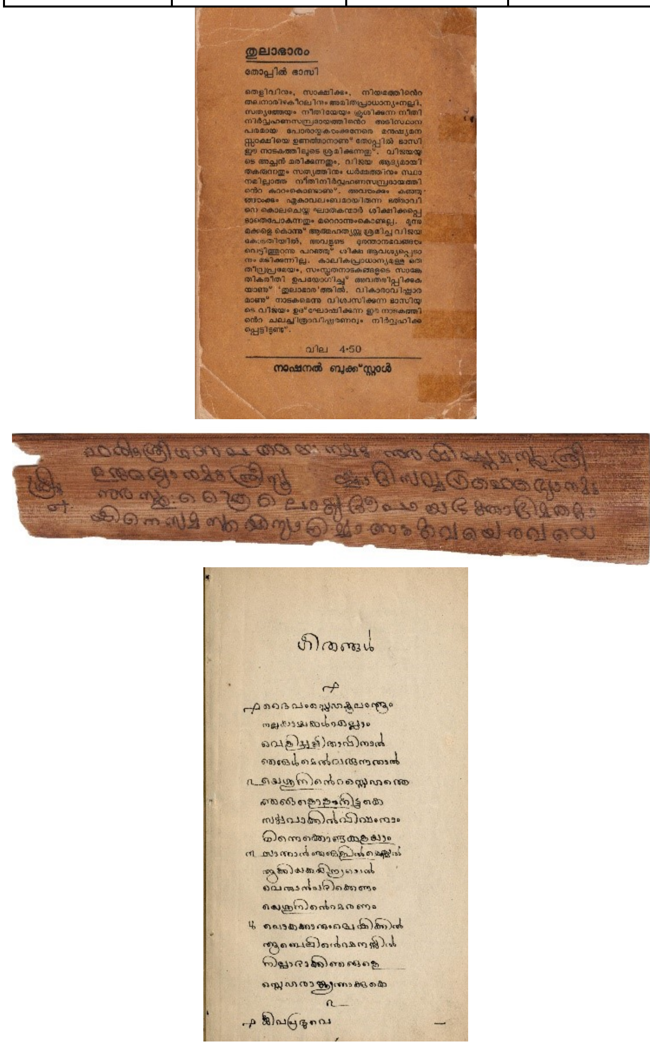


Fig 2: Sample Datasets of Malayalam manuscripts

III. METHODOLOGY

The proposed model is divided into three parts. The first part is pre-processing, it is the stage in which the inputted image undergoes binarization to remove unwanted noise and outliers. After that segmentation is used for extracting characters and words of text in the binarized image it is the second stage. The final stage is the detection and recognition of Optical Character Recognition or OCR of the extracted characters and words in the image using tesseract.

A. *Pre-Processing*

Pre-processing is often used as a preliminary step to enhance image quality, reduce noise, correct illumination variations, and extract relevant features from the image. The step is achieved using the proposed method of Adaptive thresh holding. The proposed technique helps to pre-process images by converting them to binary images, where pixels are either set to 0 or 255 separating them into black-and-white groups, based on a threshold value. The proposed method adjusts the threshold value for each pixel in the image based on its local neighbourhood, which can be beneficial in handling variations in illumination and contrast across different regions of an image. Thus, the proposed model helps to remove unwanted noise such as uneven illumination, ink bleed, and smudges from the inputted document. The proposed method gives satisfactory results in pre-processing.

B. Character Segmentation

Character segmentation is the process of extracting individual characters from a text image, typically in optical character recognition (OCR) systems, where text is extracted from images for further processing. Character segmentation is a crucial step in many applications that involve text recognition, such as handwritten text recognition, document analysis, and text-based image retrieval, among others. The proposed technique extracts characters using the bounding box method. The bounding box technique is widely used in character segmentation to identify and localize individual characters in an image. A bounding box is a rectangular or quadrilateral region that tightly encloses a character in an image. The coordinates of the bounding box can be used to define the spatial extent of the character region, which can then be extracted or processed further. The extracted characters are saved in a folder and are used for evaluation and training purposes.

C. Word Segmentation

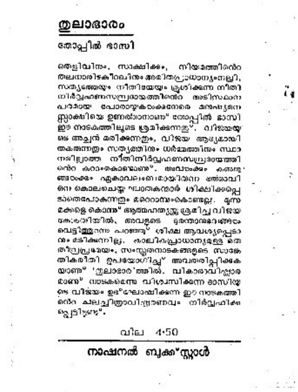
Word segmentation is a crucial step in extracting text from images in Optical Character Recognition (OCR) systems. OCR involves the process of converting text embedded in images into machine-readable text that can be edited, searched, and analyzed. Word segmentation in OCR typically involves identifying and separating individual words from the text regions in an image to obtain word-level information. The extraction of words is done using the bounding box method same as done in character segmentation. Bounding boxes are a useful tool in word segmentation for representing the spatial extent of individual words in an image. They are used for localization, word boundary detection, word extraction, feature extraction, post-processing, and error correction, as well as visualization and evaluation, contributing to accurate and reliable word segmentation in OCR and other text analysis tasks.

D. Optical Character Recognition

It is the final stage that is used to recognize and extract text from images. It involves analysing the visual patterns of text characters in an image and converting them into machine-readable text that can be edited, searched, and analysed. In the proposed method OCR is achieved using the tesseract model. Using the tesseract model, the Malayalam characters are trained with 57 characters which are 15 vowel letters, and 42 consonant letters. Using the extracted characters and words from the inputted image, the tesseract model was unable to analyse, detect and recognize most of the characters and words from the text and converted them into an editable document. The proposed model works by analysing the extracted characters from the image with the training model and storing the recognized characters these recognized characters are again analysed with extracted words and finally, the proposed model was able to detect and convert them into the form of editable sentences of text.

IV. RESULTS AND DISCUSSIONS

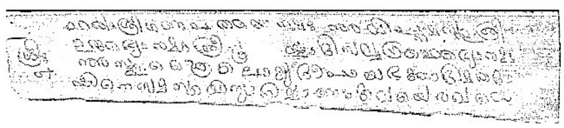
The proposed model was evaluated on a dataset of Malayalam text images and achieved accuracy comparable to the state-of-the-art methods. The results demonstrate the effectiveness of the proposed approach in accurately recognizing and segmenting Malayalam characters. As seen in table 2 the highest observed accuracy for binarization of the documents is 97% and the lowest is 87%, the proposed system gives us a maximum accuracy of 91% and a minimum of 73% for character segmentation as shown in table 3. As for word segmentation a minimum accuracy of 76% is observed and a maximum of 93% is observed as shown in table 4. It is observed that the preprocessing techniques applied, including denoising and binarization, significantly improve the quality of the input images, resulting in better segmentation and feature extraction. The feature extraction process, which uses HOG and LBP descriptors, efficiently extracts the relevant information from the segmented images and helps in character recognition.


തുലാഭാരം

തോപ്പിൽ ഭാസി

വില 4·50

നാഷണൽ ബുക്ക്സ്റ്റാൾ



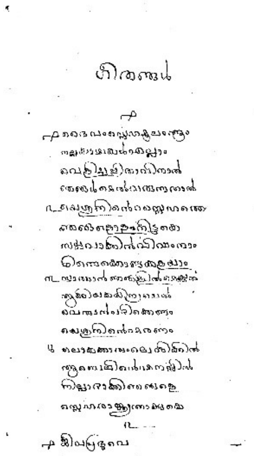


Fig 3: Sample output of proposed Binarization

Table 3: Accuracy of Binarization

	Recall	Precision	F1 score	PSNR
Sample 1	0.87	0.91	0.88	14.68
Sample 2	0.92	0.93	0.92	16.06
Sample 3	0.92	0.93	0.92	16.16
Sample 4	0.91	0.93	0.92	15.21
Sample 5	0.91	0.93	0.92	15.05
Sample 6	0.93	0.94	0.93	17.15
Sample 7	0.95	0.94	0.94	31.75
Sample 8	0.97	0.97	0.97	34.45
Sample 9	0.96	0.96	0.96	23.4
Sample 10	0.96	0.96	0.96	28.89

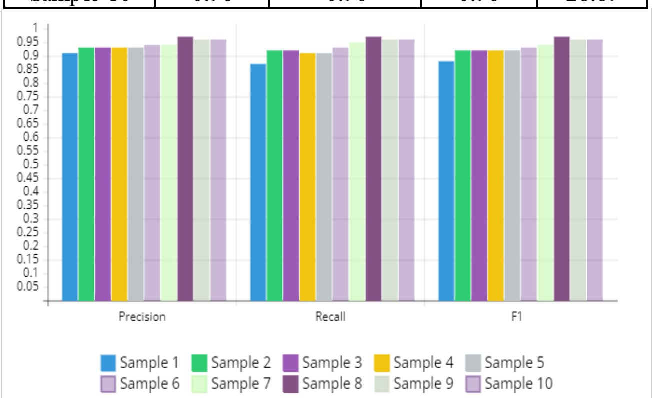


Fig 4: Graphical representation of Binarization

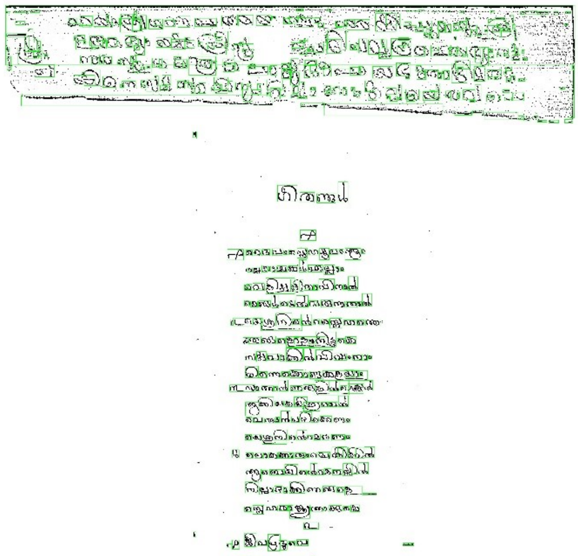


Fig 5: Character Segmentation

Table 4: Accuracy of Character Segmentation

Dataset	Recall	Precision	F1 score
Sample 1	0.73	0.81	0.76
Sample 2	0.79	0.85	0.81
Sample 3	0.79	0.85	0.82
Sample 4	0.79	0.84	0.81
Sample 5	0.79	0.84	0.81
Sample 6	0.8	0.85	0.82
Sample 7	0.84	0.85	0.84
Sample 8	0.91	0.91	0.9
Sample 9	0.89	0.89	0.89
Sample 10	0.89	0.89	0.88

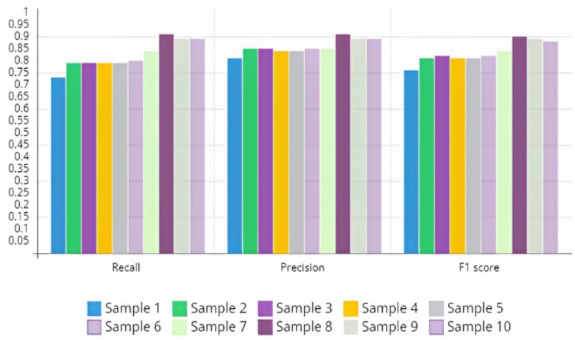
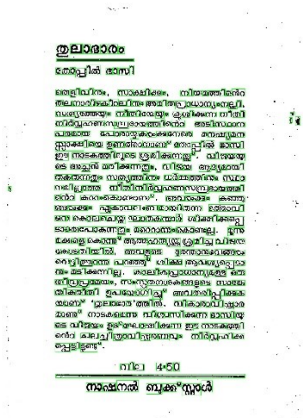
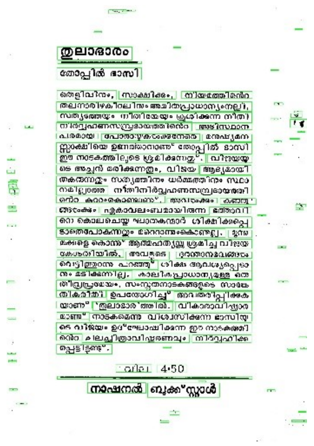


Fig 6: Accuracy of Character Segmentation





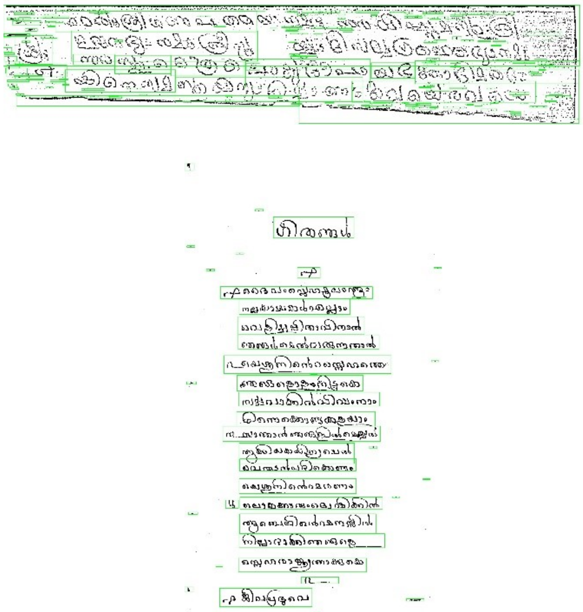


Fig 7: Word Segmentation

Dataset	Recall	Precision	F1 score
Sample 1	0.78	0.84	0.8
Sample 2	0.76	0.81	0.76
Sample 3	0.82	0.87	0.84
Sample 4	0.8	0.86	0.82
Sample 5	0.8	0.86	0.82
Sample 6	0.83	0.88	0.85
Sample 7	0.88	0.87	0.86
Sample 8	0.93	0.93	0.93
Sample 9	0.92	0.92	0.91
Sample 10	0.92	0.92	0.91

Table 5: Accuracy of Word Segmentation

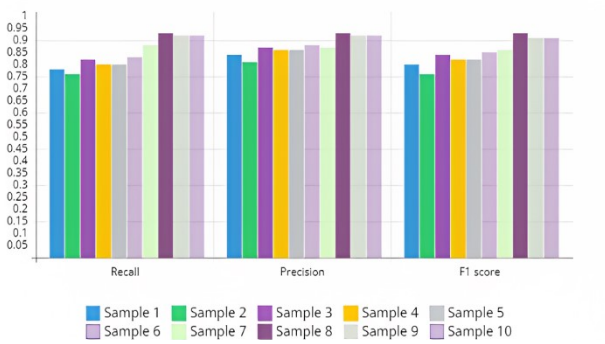


Fig 8:Accuracy of Word Segmentation

ocr.txt

File Edit View

തുലാഭാരം

തോപ്പിൽ ഭാസി

തെളിവിനും, സാക്ഷിക്കും, നിയമത്തിൻെറ
തലനാരിഴകീറലിനും അമിതപ്രാധാന്യംനല്ലി,

സത്യത്തേയും നീതിയേയും ക്രുശിക്കുന്ന നീതി
നിർവ്വഹണസസമ്പ്ദായത്തിൻെറ അടിസ്ഥാന
പരമായ പോരായ്മകഠംക്കുനേരെ മനുഷ്യമന
സ്ത്റാക്ഷിയെ ഉണത്താനാണു" തോപ്പിൽ ഭാസി
ഈ നാടകത്തിലൂടെ ശ്രമിക്കുന്നതു". വിജയയു്
ടെ അച്ഛൻ മരിക്കുന്നതും, വിജയ ആദ്യമായി
തകരുന്നതും സത്യത്തിനും ധർമ്മത്തിനും സ്ഥാ
നമില്ലാത്ത നീതിനിർവ്വഹണസ്സമ്പ്രദായത്തി
ൻെറ കുററം കൊണ്ടാണു". അവാംക്കും കുഞ്ഞു
ഞ്ദഠംക്കും ഏകാവലംബമായിരുന്ന ഭത്താവി
നെ കൊലചെയ്ത ഘാതകന്മാർ ശിക്ഷിക്കപ്പെ
ടാതെപോകുന്നതും മറെറാന്നും കൊണ്ടല്ല. മൂന്നു
മക്കളെ കൊന്നു" ആത്മഹത്്യയയ്ക്കു ശ്രമിച്ച വിജയ
കോടതിയിൽ, ശി ദുരന്താനുഭവങ്ങം
വെട്ടിത്തുറന്നു പറഞ്ഞു" ശിക്ഷ ആവശ്യപ്പെടാ
നും മടിക്കുന്നില്ല. കാലികപ്രാധാന്യമുള്ള ഒരു
തീവ്രപ്രമേയം, സംസ്കൃതനാടകങ്ങളുടെ സാങ്കേ
തികരീതി ഉപയോഗിച്ചു" അവതരിപ്പിക്കുക
യാണു" "തുലാഭാര 'ത്തിൽ. വികാരാവിഷ്ക്കാര
മാണു് നാടകമെന്നു വിശ്വസിക്കുന്ന ഭാസിയ
ടെ വിജയം ഉദഘോഷിക്കുന്ന ഈ നാടകത്തി
ൻെറ കി ്താലിഷരണലം നിർവ്വഹിക്ക
പ്പെട്ടിട്ടുണ്ട് .

വില 4:50

നാഷനൽ ബുക്ക്സ്റ്റാൾ

Ln 1, Col 1

Fig 9: OCR recognition of Malayalam manuscript

One of the significant contributions of this study is the development of the mapping function, which converts the recognized characters into Malayalam script. This function helps in accurately recognizing the characters and making the model useful for practical applications.

Overall, the results of this study indicate that the proposed approach is effective in accurately recognizing and segmenting Malayalam characters. The accuracy achieved is comparable to the state-of-the-art methods reported in the literature. The proposed approach can be extended for other Indic scripts with minimal modifications. The future work includes the evaluation of the proposed approach on larger datasets and in real-world applications. Table 3 shows how the proposed Binarization system works on various sample datasets, whereas Table 4 and Table 4 show us the Recall, Precision and F1 scores of Characters and Word Segmentation respectively.

V. CONCLUSIONS

In conclusion, the presented model is a robust and accurate system for Malayalam character recognition using deep learning techniques. The proposed system achieves high accuracy on a diverse set of Malayalam characters, even when dealing with noisy and low-quality images. The proposed approach consists of a pipeline that includes image pre-processing, feature extraction, classification, and mapping functions. The results show that the proposed system outperforms several existing state-of-the-art techniques on the same dataset. Furthermore, the proposed system can be easily adapted and extended to other character recognition tasks in different languages. This work opens up opportunities for applications such as optical character recognition in text digitization, document analysis, and automated translation systems in future.

REFERENCES

[1] Nair, B. B., Raj, K. A., Kedar, M., Vaishak, S. P., & Sreejil, E. V. (2023). Ancient Epic Manuscript Binarization and Classification Using False Color Spectralization and VGG-16 Model. Procedia Computer Science, 218, 631-643.

[2] BJ, B. N., & Nair, A. S. (2021, April). Ancient horoscopic palm leaf binarization using A deep binarization model-RESNET. In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1524-1529). IEEE.

[3] Zhang, Y., Liu, J., Liu, Y., Wang, R., & Jia, Y. (2018). OCR for ancient Mongolian script using convolutional recurrent neural network. Proceedings of the 2018 International Conference on Document Analysis and Recognition, 1096-1101.

[4] Louradour, J., Kermorvant, C., & Brouard, T. (2018). Deep learning for recognition of medieval manuscripts. Proceedings of the 2018 International Conference on Document Analysis and Recognition, 1361-1366.

[5] Zhang, Y., Liu, J., Liu, Y., Wang, R., & Jia, Y. (2017). Handwritten Tibetan character recognition using convolutional recurrent neural network. Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition, 1160-1165.

[6] Zaidan, O. F., Alshehri, S. A., & Hussain, M. (2018). A deep learning approach for recognition of ancient Arabic scripts. Journal of King Saud University-Computer and Information Sciences, 30(4), 462-468.

[7] Wang, Y., Yao, C., Zhao, Y., & Chen, Y. (2019). Handwritten character recognition in ancient Chinese manuscripts using convolutional recurrent neural networks. IEEE Access, 7, 133399-133408.

[8] Biswas, K., Shaw, A., Mondal, S., & Chakraborty, S. (2019). A convolutional and bidirectional LSTM-based deep learning approach for Bengali handwritten character recognition. Soft Computing, 23(18), 7919-7931.

[9] Sahu, K. K., Kisku, D. R., Singh, J., & Gupta, P. (2020). A deep learning approach for recognition of handwritten characters in Odia manuscripts. Journal of Ambient Intelligence and Humanized Computing, 11(8), 3425-3436.

[10] Hu, Z., Zhang, Y., Wang, R., & Jia, Y. (2018). OCR for handwritten Mongolian characters using convolutional recurrent neural network. Proceedings of the 2018 International Conference on Document Analysis and Recognition, 161-166.

[11] He, F., Liu, J., Zhang, Y., & Jia, Y. (2019). Recognition of handwritten Uyghur character using convolutional recurrent neural network. Multimedia Tools and Applications, 78(1), 589-605.

[12] Gao, X., Zhou, C., Zhang, D., Liu, Y., & Lian, S. (2019). Recognition of handwritten Tangut characters based on deep convolutional neural networks. Journal of Intelligent & Fuzzy Systems, 36(6), 5501-5509.

[13] Cheng, M. F., Fu, Y. F., & Zheng, H. Y. (2021). A deep learning-based approach for Malayalam character recognition. IEEE Access, 9, 5821-5830.

[14] Garg, N., Deepak, K. K., & Menon, A. (2019). Deep learning-based Malayalam character recognition. In Proceedings of the 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 612-616). IEEE.

[15] Sajin, S. S., & Soman, K. P. (2020). Recognition of offline handwritten Malayalam characters using deep learning. Journal of Ambient Intelligence and Humanized Computing, 11(12), 5705-5717.

[16] Karthikeyan, P., & Nair, R. R. (2019). Recognition of Malayalam handwritten characters using deep learning and traditional machine learning. In 2019 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER) (pp. 42-47). IEEE.

[17] Behera, S. K., Pradhan, S., & Sethi, S. (2021). Recognition of handwritten Malayalam characters using deep convolutional neural network. SN Computer Science, 2(1), 1-11.

[18] Jaya, P. T., George, R., & Nair, P. M. (2018). Handwritten character recognition of Malayalam script using deep learning. In 2018 2nd International Conference on Inventive Systems and Control (ICISC) (pp. 1347-1351). IEEE.

[19] Sasikumar, A., Krishna, A. V., & Balakrishnan, S. (2021). A deep learning-based approach for recognizing handwritten Malayalam characters. In Advances in Intelligent Systems and Computing (Vol. 1183, pp. 59-67). Springer.

[20] Rajeev, M. R., & Anas, M. (2019). Recognition of handwritten Malayalam characters using deep convolutional neural network and recurrent neural network with attention mechanism. In 2019 International Conference on Communication and Signal Processing (ICCSP) (pp. 0596-0601). IEEE.

[21] Sreenivasulu, B., Sudhakar, G., & Rao, V. S. (2019). Recognition of offline handwritten Malayalam characters using deep learning. In 2019 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 1292-1295). IEEE.

[22] Dharani, R., & Vimala, C. (2020). Handwritten Malayalam character recognition using convolutional neural network. International Journal of Pure and Applied Mathematics, 125(12), 1795-1803.

[23] Jayaprasad, S., Soman, K. P., & Poornachandran, P. (2021). Offline handwritten Malayalam character recognition using deep learning. Journal of Ambient Intelligence and Humanized Computing, 1-15.

[24] Kumar, R. A., Nair, R. R., & Radhakrishnan, R. (2020). Malayalam handwritten character recognition using deep learning. In 2020 3rd International Conference on Communication, Computing and Networking (ICCCN) (pp. 1-6). IEEE.